RECEIVED
CENTRAL FAX CENTER
NOV 0 8 2005

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

BOARD OF PATENT APPEALS AND INTERFERENCES

In re patent application of:

Kreulen, et al.

Serial No.: 09/629,831

Filed: July 31, 2000

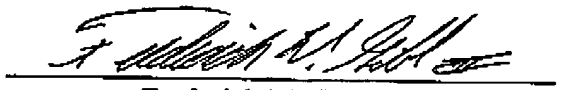
Group Art Unit: 2176

Examiner: Smith, Peter J.

Atty. Docket No.: AM9-99-0157

Certificate of Transmission by Facsimile

I hereby certify that this correspondence is being facsimile transmitted to the United States Patent and Trademark Office (Fax No. 571-273-8300) on 11/8/05



Frederick W. Gibb, III

For: METHOD FOR GENERATION OF AN N-WORK PHRASE DICTIONARY FROM A TEXT CORPUS

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

APPELLANTS' APPEAL BRIEF

Sirs:

Appellant respectfully appeals the final rejection of claims 1-17 in the Office Action dated June 14, 2005. A Notice of Appeal was filed on September 16, 2005.

I. REAL PARTY IN INTEREST

The real party in interest is International Business Machines Corp., Armonk, New York, assignee of 100% interest of the above-referenced patent application.

II. RELATED APPEALS AND INTERFERENCES

There are no other appeals or interferences known to Appellants, Appellants' legal representative or Assignee which would directly affect or be directly affected by or have a bearing on the Board's decision in this appeal.

III. STATUS OF CLAIMS

Claims 1, 6, and 11 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Kostoff et al., hereinafter "Kostoff" (U.S. Patent No. 5,440,481). Claims 2-5, 7-10, and 12-17 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Kostoff and in further view of Kirsch et al., hereinafter "Kirsch"(U.S. Patent No. 6,070,158), Kobayashi (U.S. Patent No. 5,742,834) and Turney (U.S. Patent No. 6,470,307).

IV. STATUS OF AMENDMENTS

An After-final Amendment was filed on August 11, 2005. An Advisory Action dated August 23, 2005 indicated that, upon filing an appeal, the Amendment filed on August 11, 2005 did not place the application in condition for allowance, and that the rejections of claims would remain. The claims shown in the appendix are shown in their amended form as of the April 4, 2005 Amendment.

V. SUMMARY OF CLAIMED SUBJECT MATER

By using a "maximum dictionary size" as a vehicle to control how many terms are to be used in a phrase search (e.g., limiting the size of the dictionary before the frequency of phrases in the document that contain words in the dictionary is determined), the

invention provides an automated methodology which, without additional user input, reduces the size of the data that must be processed, thereby making the processing more efficient and conserving precious processing resources.

As described on page 16, line 15-page 17, line 5 of the application, some benefits which flow from this invention are derived from the ability to readily adapt the creation of text dictionaries containing both words and phrases to the capabilities of the computer hardware available. The invention allows the user to specify the dictionary size up front, without reference to the size or complexity of the data set to be analyzed, and the invention returns all of the most frequent terms which can fit within this memory constraint. This allows the user to analyze text data sets of arbitrary size and complexity on computer hardware of fixed memory and computational speed. Creation of word/phrase dictionaries on text data sets further allows for the analysis of unstructured text information in a semi-structured manner. Data mining algorithms and statistical measure can now be applied to the data to discover interesting relationships and trends. Dictionary creation is thus the first critical step in data mining and analysis of text data sets. Being able to generate such dictionaries quickly and efficiently and with high quality is therefore of key importance to successful text mining.

Referring to Figure 1, the invention performs a "first pass" (independent claim 6) on the set of text documents, as shown in the item 10. Next, in item 11, the invention creates a Hashtable and keeps only the most frequently occurring words in the Hashtable. Thus, the invention finds the V most frequently occurring words in the word-count Hashtable and conserves memory by removing from the Hashtable all words that occur with less frequency than the V most frequently occurring words. This is defined in the independent claims as "determining a frequency of each word in each of said documents; creating a dictionary of most frequently occurring words in said documents as limited by said maximum dictionary size, such that said dictionary contains less than all words in said documents."

Then, as shown in item 12, the invention performs a "second pass" (independent claim 6) on the input set of text documents. In item 13, the invention adds phrases that are made up only of words in the word-count Hashtable to a phrase-count Hashtable. Finally, in item 14, the invention finds the most frequently occurring V words and phrases in the Hashtable and creates a dictionary of words and phrases from the Hashtable. This is defined in claims as "adding most frequently occurring phrases to said dictionary; and outputting said most frequently occurring words and said most frequently occurring phrases as said dictionary, wherein said dictionary size limits the number of words and phrases maintained in said dictionary."

As described on page 15, lines 1-9 of the application, previous methods for generating a dictionary from a text corpus focused on individual words only or have generated phrases based on a linguistic analysis. The invention's methodology is purely lexical in nature and thus generalizes to multiple languages and to ungrammatical text. Previous methodologies have suggested a lexical phrase generation technique and have not described the space and time efficient implementation for discovering such phrases that the invention utilizes. The invention's implementation is designed to quickly find a maximal frequency term dictionary of a given size using the smallest possible amount of memory.

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

The issues presented for review are whether claims 1, 6, and 11 are unpatentable under 35 U.S.C. §103(a) as being unpatentable over Kostoff et al., hereinafter "Kostoff" (U.S. Patent No. 5,440,481). Claims 2-5, 7-10, and 12-17 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Kostoff and in further view of Kirsch et al., hereinafter "Kirsch"(U.S. Patent No. 6,070,158), Kobayashi (U.S. Patent No. 5,742,834) and Turney (U.S. Patent No. 6,470,307).

VII. ARGUMENT

A. The Rejection Based on Kostoff

1. The Position in the Office Action

The Office Action states:

> Regarding independent claim 1, Kostoff teaches determining a frequency of each word in each document in fig. 2, table 1, col. 4 lines 50-68, and col. 6 line 65 - col, 7 line 11. Kostoff teaches creating a table of most frequently occurring words in the documents in fig. 2, table 1, col. 4 lines 50-68, and col, 6 line 65 col. 7 line 11. Kostoff teaches determining a frequency of phrases in each document that could contain only words in a table in fig. 2, table 1, col. 4 lines 50-68, and col, 6 line 65 - col. 7 line 11. Kostoff teaches outputting the most frequently occurring words and most frequently occurring phrases as a dictionary in fig. 2 and col. 4 lines 64-68.
>
> Kostoff does not specifically teach inputting a maximum dictionary size and limiting the dictionary to the inputted maximum dictionary size, such that the dictionary contains less than all words in the documents. However, Kostoff does acknowledge the importance and limitation of memory size for storing a list of trivial words in col. 4 lines 44-45, This list is a precursor to the dictionary, however it teaches one of ordinary skill in the art at the time of the invention the relevance of memory storage size. Kostoff also teaches selecting a portion of the word and phrase dictionary in col. 5 line 59- col. 6 line 64. Kostoff uses an example of selecting the 60 most often repeated phrases. Kostoff notes that more or less

than 60 most often repeated phrases may be selected at the discretion of the user.

In light of these teachings of Kostoff, one of ordinary skill in the art at the time of the invention would have truncated the dictionary of Kostoff at the user inputted number of most often repeated phrases in the event the dictionary had to reside within a limited memory storage. The teaching of Kostoff of possible memory storage constraints having an impact on a list size in col. 4 lines 44-45 would have motivated and taught insight to the person of ordinary skill in the art at the time of the invention to have made this modification It would have been obvious to one of ordinary skill in the art at the time of the invention to have discarded the less frequent terms below the population threshold inputted by the user because they would not have been of further use in determining the themes of the text to prepare it for clustering with other documents, Eliminating the unused terms would have desirably saved memory as seen in col. 4 lines 44-45, Only the top set of words and phrases determined by the user would have been used and therefore it would have been obvious to have only retained those words and phrases in the dictionary.

Regarding independent claim 6, Kostoff teaches determining a frequency of each word in each document in fig. 2, table I, col. 4 lines 50-68, and col. 6 line 65- col, 7 line 11. Kostoff teaches creating a table of most frequently occurring words in the documents in fig. 2, table 1, col, 4 lines 50-68, and col. 6 line 65- col. 7 line 11. Kostoff teaches determining a frequency of phrases in each document that could contain only words in a table in fig. 2, table 1, col. 4 lines 50-68, and col. 6 line 65 - col. 7 line 11. Kostoff teaches outputting the most frequently occurring words and most frequently occurring phrases as a dictionary in fig. 2 and col. 4 lines 64-68.

Kostoff does not specifically teach inputting a maximum dictionary size and limiting the dictionary to the inputted maximum dictionary size, such that the dictionary contains less than all words in the documents. However, Kostoff does acknowledge the importance and limitation of memory size for storing a list of trivial words in col. 4 lines 44-45. This list is a precursor to the dictionary, however it teaches one of ordinary skill in the art at the time of the invention the relevance of memory storage size. Kostoff also teaches selecting a portion of the word and phrase dictionary in col. 5 line 59- col. 6 line 64. Kostoff uses an example of selecting the 60 most often repeated phrases. Kostoff notes that more or less than 60 most often repeated phrases may be selected at the discretion of the user.

In light of these teachings of Kostoff, one of ordinary skill in the art at the time of the invention would have truncated the dictionary of Kostoff at the user inputted number of most often repeated phrases in the event the dictionary had to reside within a limited memory storage. The teaching of Kostoff of possible memory storage constraints having an impact on a list size in col. 4 lines 44-45 would have motivated and taught insight to the person of ordinary skill in the art at the time of the invention to have made this modification, It would have been obvious to one of ordinary skill in the art at the time of the invention to have discarded the less frequent terms below the population threshold inputted by the user because they would not have been of further use in determining the themes of the text to prepare it for clustering with other documents, eliminating the unused terms would have desirably saved memory as seen in col. 4 lines 44-45. Only the top set of words and phrases determined by the user would have been used and therefore it would have been obvious to have only retained those words and phrases in the dictionary, Kostoff does not explicitly teach the creation of the word and phrases lists in two separate passes

through the document. One of ordinary skill in the art at the time of the invention would have known how to create the two lists in separate passes through the document. It would have been obvious to one of ordinary skill in the art at the time the invention was made to use their skill in the art to have created each list as a result of each of two passes through the document. This would have been obvious and necessary in order to create the second list since the phrase selection would have been dependent on the contents of the first list.

Regarding independent claim 11, Kostoff teaches determining a frequency of each word in each document in fig. 2, table 1, col. 4 lines 50-68, and col. 6 line 65- col. 7 line 11. Kostoff teaches creating a table of most frequently occurring words in the documents in fig. 2, table 1, col. 4 lines 50-68, and col. 6 line 65 - col. 7 line 11. Kostoff teaches determining a frequency of phrases in each document that could contain only words in a table in fig. 2, table 1, col. 4 lines 50-68, and col. 6 line 65 - col. 7 line 11. Kostoff teaches outputting the most frequently occurring words and most frequently occurring phrases as a dictionary in fig. 2 and col. 4 lines 64-68, Kostoff does not specifically teach inputting a maximum dictionary size and limiting the dictionary to the inputted maximum dictionary size, such that the dictionary contains less than all words in the documents, However, Kostoff does acknowledge the importance and limitation of memory size for storing a list of trivial words in col. 4 lines 44-45, This list is a precursor to the dictionary, however it teaches one of ordinary skill in the art at the time of the invention the relevance of memory storage size. Kostoff also teaches selecting a portion of the word and phrase dictionary in col, 5 line 59- col. 6 line 64. Kostoff uses an example of selecting the 60 most often repeated phrases. Kostoff notes that more or less than 60 most often repeated phrases may be selected at the discretion of the user.

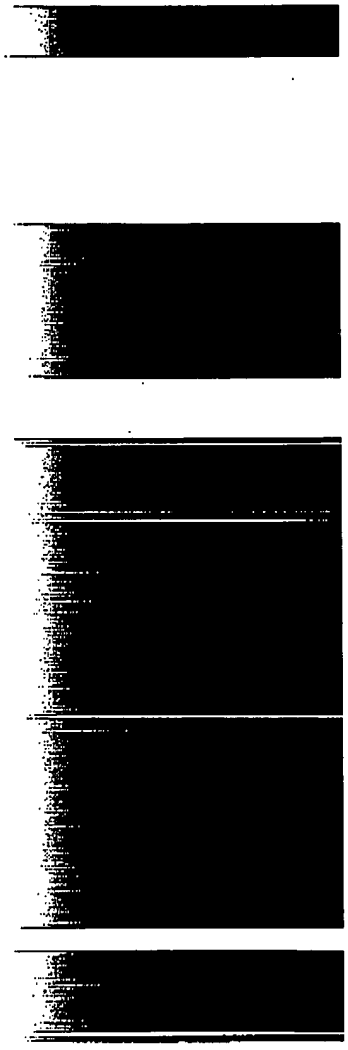
In light of these teachings of Kostoff, one of ordinary skill in the art at the time of the invention would have truncated the dictionary of Kostoff at the user inputted number of most often repeated phrases in the event the dictionary had to reside within a limited memory storage. The teaching of Kostoff of possible memory storage constraints having an impact on a list size in col. 4 lines 44-45 would have motivated and taught insight to the person of ordinary skill in the art at the time of the invention to have made this modification It would have been obvious to one of ordinary skill in the art at the time of the invention to have discarded the less frequent terms below the population threshold inputted by the user because they would not have been of further use in determining the themes of the text to prepare it for clustering with other documents. Eliminating the unused terms would have desirably saved memory as seen in col. 4 lines 44-45, only the top set of words and phrases determined by the user would have been used and therefore it would have been obvious to have only retained those words and phrases in the dictionary.

Response to Arguments

Appellants arguments filed 4/4/2005 have been fully considered but they are not persuasive, Regarding Appellant's arguments in pages 7-10 that the invention as presented in independent claims 1, 6, and 11 is not obvious over Kostoff et al. (hereinafter "Kostoff"), the Examiner respectfully disagrees. The Examiner admits Kostoff does not directly anticipate the claimed invention, However, the Examiner believes the teachings of Kostoff in col, 4 lines 39-49 are important in that this teaching would have enlightened one of ordinary skill in the art at the time of the invention to have modified Kostoff to have created the claimed invention, Appellant's invention limits the dictionary to the most frequently occurring words as limited by the

maximum dictionary size. All of the other words are discarded from use in the dictionary. Kostoff teaches that a trivial phrase list is preferably applied prior to or during processing the text such that any word or phrase contained in the trivial phrase list is not included in the dictionary. Kostoff teaches that the list of trivial phrases may be any words that the user wishes to have included in the list and also that the list may be unlimited in size. Because the trivial phrase list of Kostoff may be unlimited in size to the user's liking, the Examiner believes that Kostoff teaches that the trivial phrase list does not necessarily only contain words meaningless to document content such as "to" and "if', but rather may also contain words and phrases the user deems not important. Thus, in the context and terminology of Kostoff the Examiner believes Appellant's invention essentially makes any word below a certain keyword threshold frequency (determined by a maximum dictionary size) a trivial word to be excluded from the dictionary. The Examiner believes that if the trivial word list inputted to modify the dictionary prior to its creation contains all the words below a certain frequency threshold, then Kostoff would produce the same dictionary as that of the claimed invention.

In response to Appellant's point on page 8 that Kostoff states in col, 4 lines 52-55 that the system and methodology are required to use the entire full-text database to create lists and phrases, the Examiner notes that this step occurs after the trivial phrase list is excluded from processing and entry into the dictionary. Thus, the "entire full-text" mentioned in the cited section of Kostoff is not really the entire full-text, but rather the entire full-text minus the trivial phrase list, Therefore, the Examiner does not believe this is evidence that Kostoff teaches away from Appellant's claimed invention. The Examiner does not agree with the distinction presented on pages 9 and 10 of Appellant's response because Kostoff does not maintain a list of all potential phrases in the text



corpus because in Kostoff the phrases deemed trivial by the user are not entered into the dictionary. The Examiner believes Kostoff suggests to one of ordinary skill in the art at the time of the invention reasons to modify Kostoff to have created the invention as presented in independent claims 1, 6, and 11.

2. Appellants' Position

a. Independent Claims 1 and 11

The Office Action accurately states (on pages 14-15) that the claimed invention limits the dictionary to the most frequently occurring terms, as limited by the preset "maximum dictionary size". Then, the claimed invention can search the associated document for phrases that contain only these terms and produce a dictionary of most frequently occurring phases and terms. By using the "maximum dictionary size" as the vehicle to control how many terms are to be used in the phrase search (e.g., limiting the size of the dictionary before the frequency of phrases in the document that contain words in the dictionary is determined), the invention provides an automated methodology which, without additional user input, reduces the size of the data that must be processed.

The June 14, 2005, Office Action argues (on pages 14-15) that because Kostoff removes a manually created trivial phrase list from the dictionary before using the dictionary to search for phrases in the associated documents, one ordinarily skilled in the art would be motivated to take efforts to reduce the dictionary size before searching for phrases, as in the claimed invention.

In other words, the Office Action presents an argument that, by limiting the dictionary to only the most frequently occurring words (as limited by the "maximum dictionary size"), the claimed invention essentially removes all "trivial" words from the dictionary before searching for phrases. Since Kostoff also teaches that all trivial words ("to", "if", etc.) should be removed from the dictionary before searching for phrases, the

Office Action argues that Kostoff would have suggested the claimed invention to one ordinarily skilled in the art.

While this argument is initially appealing, it is Appellants' position that Kostoff does not teach one ordinarily skilled in the art to limit which words can be added to the dictionary according to the "maximum dictionary size". Independent claims 1 and 11 provide for "creating a dictionary of most frequently occurring words in said documents as limited by said maximum dictionary size." Therefore, with the invention, the decision of which words to include in, or exclude from the dictionary is determined just by entering the "maximum dictionary size". To the contrary, with Kostoff the manually created list of "trivial" words that are excluded from the dictionary is used to limit which words are excluded from the dictionary (col. 4, lines 39-42).

Contrary to the highly manual process described in Kostoff, the claimed methodology is fully automated (the only input required being the "maximum dictionary size", which can simply be equal to the available memory or manually preset by the user), while Kostoff requires the user to manually create the trivial phrase list (col. 4, lines 39-42). The efficiency gains of the automated inventive methodology when compared to the manual system described in Kostoff are substantial.

Further, the removal of trivial words ("to", "if", etc.) in Kostoff is actually more similar to the claimed removal of a manually created list of "stop" words (the, and, a, there, is, than) as defined by dependent claims 2-3, 7-8, and 12-13. The rules of claim differentiation and construction provide that each claim in a patent is presumptively different in scope. Therefore, the removal of trivial stop words in the dependent claims is different that the removal of words based on the maximum dictionary size in the independent claims. Here, the removal of a manually created list of trivial phrases ("to", "if", etc.) in Kostoff is equivalent to the claimed removal of a manually created list of stop words (the, and, a, there, is, than). Thus, the claimed method of limiting the dictionary according to a maximum size is a distinct feature from the removal of trivial or stop words and phrases. Therefore, it is Appellants' position that the discussion in Kostoff regarding the list of trivial words and phrases teaches no more that what is

performed when the claimed invention removes stop words. There is nothing within Kostoff which would suggest that this removal of trivial or stop words would lead one ordinarily skilled in the art to limit which words are to be included in the dictionary according to a "maximum dictionary size".

The creation of a manual list of trivial words ("to", "if", etc.) and its removal from the dictionary does not suggest the claimed automated methodology which simply and automatically limits the dictionary using a size limit. It is Appellants' position that the requirement that a manually created list be used to limit the dictionary size teaches away from the claimed automated methodology which does not require the user to specify any words, but instead merely eliminates the least frequent words from the dictionary. Further, the claimed invention may actually include all "trivial" words (if these stop words are not otherwise removed as provided in the dependent claims) as these words may be the most common. Again, the claimed invention removes the "most frequently occurring words in said documents as limited by said "maximum dictionary size"" and trivial or stop words may actually be the most common (if otherwise not removed in a separate processing step).

One difference between the claimed invention and Kostoff is that the size of the dictionary is limited before the frequency of phrases in the document that contain words in the dictionary is determined. This is important because the number of phrases grows exponentially with the size of the corpus. Simply removing a list of trivial phrases may not reduce the dictionary size (especially if the manually created list of trivial phrases finds no matches in the dictionary). By reducing the size of the dictionary before determining the frequency of phrases containing words in the dictionary, the claimed invention produces exponential gains in processing speed and memory usage.

In other words, the claimed invention involves more than just reducing the dictionary to meet a memory constraint. In the claimed invention, the dictionary is reduced at a point in the processing that allows the method to substantially simplify the subsequent process of determining the frequency of phrases in the document containing words in the dictionary.

The claimed invention first limits the dictionary to only the top number of most frequently occurring words and then "after creating said dictionary" (claims 1 and 11) only considers phrases that contain these words. The invention avoids maintaining a list of all potential phrases in the text corpus. The problem with maintaining all potential phrases is that the number of phrases grows exponentially with the size of the corpus. The invention avoids this problem by fixing the size of the dictionary up front (user specified "maximum dictionary size", M), then finding the M most frequent words and then only creating phrases using these M most frequent words. To the contrary, the Kostoff patent creates a list of potentially all words and N-word phrases sorted by frequency. This is not practical for a large text corpus since such a list would be too large for most computer memory to hold.

The Office Action admits that Kostoff does not explicitly teach the claimed process of limiting the number of words that are used to establish the most frequently occurring phrases by limiting the dictionary size, but the Office Action argues that such a feature would have been obvious. More specifically, the Office Action notes that Kostoff describes that the size of the list of trivial phrases is limited by memory constraints (col. 4, lines 42-45) and that the number of phrases output to the user can be limited to those having high user interest, such as the top 60 most frequent phrases (col. 5, line 59-col. 6, line 64). Then, the Office Action argues that this would motivate one to limit the dictionary size to accommodate for hardware memory constraints.

Appellants respectfully disagree with this logical argument of obviousness for a number of reasons, including the fact that Kostoff requires that the dictionary must include all words in the documents (except for the trivial phrases mentioned above). More specifically, Figure 2 and col. 4, lines 52-55 state that the system and methodology in Kostoff "is required to use the entire full-text database to create lists of phrases." Therefore, Appellants submit that Kostoff directly teaches away from the claimed limitation that explicitly does not use all the words from the documents, and instead limits the dictionary to only the number of most frequently occurring words that will fit

into the limited size dictionary. When a reference teaches away from the claimed invention it actually demonstrates that the claimed invention is not obvious.

Thus, in a first respect, since Kostoff "is required to use the entire full-text database to create lists of phrases" it cannot teach or suggest "creating a dictionary of most frequently occurring words in said documents as limited by said "maximum dictionary size", such that said dictionary contains less than all words in said documents" as defined by independent claims 1 and 11. This requirement in Kostoff teaches away from the claimed invention and, therefore, Kostoff cannot teach or suggest this feature.

Further, the manner in which Kostoff would deal with memory and other limitations is conceptually different than the claimed invention. For example, in order to deal with memory constraints, Kostoff creates a list of trivial phrases that can be excluded from analysis (col 4, lines 39-49). This is essentially a fixed list in Kostoff that may or may not be effective in limiting the memory usage. To the contrary, the claimed invention limits the size of the dictionary, thereby providing for a more consistent and precise control of memory usage. In addition, the processing in Kostoff always uses all words in the database (except trivial words) and merely limits the number of phrases that are output (col. 5, line 59-col. 6, line 64). Thus, since all words are used in the most frequent phrase processing of Kostoff, no memory is conserved. To the contrary, the claimed invention first limits the dictionary to only the top number of most frequently occurring words and then only considers phrases that contain these words.

Therefore, it is Appellants' position that Kostoff does not teach or suggest "creating a dictionary of most frequently occurring words in said documents as limited by said "maximum dictionary size", such that said dictionary contains less than all words in said documents . . . wherein said dictionary size limits the number of words and phrases maintained in said dictionary" as defined by independent claims 1 and 11. Previous methodologies that have suggested a lexical phrase generation technique have not described the space and time efficient implementation for discovering such phrases that the invention utilizes. The invention's implementation is designed to quickly find a maximal frequency term dictionary of a given size using the smallest possible amount of

memory. Therefore, because the prior art of record does not teach or suggest the claimed invention, Appellants respectfully submit that independent claims 1 and 11 is patentable over the prior art of record.

In view the foregoing, the Board is respectfully requested to reconsider and withdraw this rejection.

b. Independent Claim 6

As shown above, Kostoff does not teach or suggest "creating a dictionary of most frequently occurring words in said documents as limited by said "maximum dictionary size"" but instead only teaches removing a manually created list of trivial words and phrases. Independent claim 6 similarly defines using the "maximum dictionary size" as the vehicle to control how many terms are to be used in the phrase search (e.g., limiting the size of the dictionary before the frequency of phrases in the document that contain words in the dictionary is determined) and is therefore not taught or suggested by Kostoff. In addition, independent claim 6 defines that such a process is performed in multiple passes and such multi-pass processing is not taught or suggested by Kostoff. The Office Action admits that Kostoff does not disclose such multi-pass processing; however, the Office Action presents an unsupported argument that such would have been obvious.

More specifically, the Office Action states that "Kostoff does not explicitly teach the creation of the word and phrases lists in two separate passes." However, the Office Action argues that "One of ordinary skill in the art at the time of the invention would have known how to create the two lists in separate passes through the document. It would have been obvious to one of ordinary skill in the art at the time the invention was made to use their skill in the art to have created each list as a result of each of two passes through the document. This would have been obvious and necessary in order to create the second list since the phrase selection would have been dependent on the contents of the first list." Appellants respectfully submit that such a position is unsupported by teachings in Kostoff

or other prior art references of record. During examination, the examiner bears the initial burden of establishing a prima facie case of obviousness. Oetiker, 977 F.2d at 1445. The prima facie case is a procedural tool, and requires that the examiner initially produce evidence sufficient to support a ruling of obviousness. Piasecki, 745 F.2d at 1475. Simply stating that a feature would have been obvious does not meet this initial burden.

Therefore, in addition to Kostoff not teaching using the "maximum dictionary size" as the vehicle to control how many terms are to be used in the phrase search (e.g., limiting the size of the dictionary before the frequency of phrases in the document that contain words in the dictionary is determined), the Office Action does not present evidence as to why it would have been obvious to perform such a process in multiple passes. Therefore, because the prior art of record does not teach or suggest the claimed invention, and because no evidence has been set forth as to why such multi-pass processing would have been obvious, Appellants respectfully submit that independent claim 6 is also patentable over the prior art of record.

In view the foregoing, the Board is respectfully requested to reconsider and withdraw this rejection.

B. The Rejection Based on Kostoff in view of Kirsch and further in view of Kobayashi and Turney

1. The Position in the Office Action

The Office Action states:

> Regarding dependent claim 2, Kostoff teaches adding words to a dictionary table in fig. 2, table 1, col. 4 lines 50-68, and col. 6 line 65 - col. 7 line 11, Kostoff teaches determining the frequency of each word remaining n the table in fig. 2, table 1, col. 4 lines 50-68, and col. 6 line
> 65 - col. 7 line 11. Kostoff teaches removing words below a frequency level from the dictionary table in col. 6 lines 2-64.

Kostoff does not teach removing punctuation and case from the documents, Kostoff does not teach removing stop words from the document. Kostoff does not teach replacing words in the documents with synonyms, Kostoff does not teach removing duplicate words from the documents, Kirsch teaches removing punctuation and case from the documents in col.12 lines 5-7. Kirsch teaches removing stop words from the document in col. 12 lines 13-15, Kobayashi teaches replacing words in the documents with synonyms in fig. 3, 34-35, and col. 1 line 54 - col. 2 line 13. Turney teaches removing duplicate words from the documents in col, 5 lines 37-38.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to have combined Kirsch, Kobayashi, and Turney into Kostoff to have created the claimed invention, It would have been obvious and desirable to have combined the punctuation and stop word removal technique of Kirsch into Kostoff so that the documents passes would have been more efficient, it would have been obvious and desirable to have combined the synonym word replacement of Kobayashi into Kostoff so that the word counts could have been uniform across all of the documents, which would have yielded the most accurate clustering results, It would have been obvious and desirable to have combined the duplicate word removal of Turney into Kostoff so that the lists would have been uniform among all the documents in the cluster, This would have yielded the most accurate clustering results among the documents.

Regarding dependent claim 3, Kostoff teaches inputting one or more stop words, synonyms and a frequency level in col. 4 lines 39-49, col. 5 lines 59-64, and col. 6 lines 60-64.

Regarding dependent claim 4, Kostoff teaches adding words to a table in fig. 2, table 1, col. 4 lines 50-68, and col, 6 line 65 - col. 7 line 11. Kostoff teaches determining the frequency of each word remaining n the table in fig. 2, table 1, col. 4 lines 50-68, and col. 6 line 65 - col. 7, line 11. Kostoff teaches removing words below a frequency level from the table in col. 6 lines 2- 64.

Kostoff does not teach removing punctuation and case from the documents, Kostoff does not teach removing stop words from the document, Kostoff does not teach replacing words in the documents with synonyms, Kostoff does not teach removing duplicate words from the documents, Kirsch teaches removing punctuation and case from the documents in col. 12 lines 5-7. Kirsch teaches removing stop words from the document in col, 12 lines 13-15. Kobayashi teaches replacing words in the documents with synonyms in fig. 3, 34-35, and col, 1 line 54 - col. 2 line 13. Turney teaches removing duplicate words from the documents in col. 5 lines 37-38.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to have combined Kirsch, Kobayashi, and Turney into Kostoff to have created the claimed invention, It would have been obvious and desirable to have combined the punctuation and stop word removal technique of Kirsch into Kostoff so that the documents passes would have been more efficient, It would have been obvious and desirable to have combined the synonym word replacement of Kobayashi into Kostoff so that the word counts could have been uniform across all of the documents, which would have yielded the most accurate clustering results, it would have been obvious and desirable to have combined the duplicate word removal of Turney into Kostoff so that the lists would have been uniform among all the documents in the cluster, This would have

yielded the most accurate clustering results among the documents.

Regarding dependent claim 5, Kostoff teaches inputting one or more stop words, synonyms and a frequency level in col. 4 lines 39-49, col. 5 lines 59-64, and col. 6 lines 60-64.

Regarding dependent claim 7, Kostoff teaches adding words to a dictionary table in fig. 2, table 1, col. 4 lines 50-68, and col. 6 line 65- col. 7 line 11. Kostoff teaches determining the frequency of each word remaining n the table in fig. 2, table 1, col. 4 lines 50-68, and col. 6 line
65 col, 7 line 11. Kostoff teaches removing words below a frequency level from the dictionary table in col, 6 lines 2-64.

Kostoff does not teach removing punctuation and case from the documents, Kostoff does not teach removing stop words from the document, Kostoff does not teach replacing words in the documents with synonyms, Kostoff does not teach removing duplicate words from the documents, Kirsch teaches removing punctuation and case from the documents in col. 12 lines 5-7. Kirsch teaches removing stop words from the document in col. 12 lines 13-15. Kobayashi teaches replacing words in the documents with synonyms in fig. 3, 34-35, and col. 1 line 54 - col. 2 line 13. Turney teaches removing duplicate words from the documents in col, 5 lines 37-38.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to have combined Kirsch, Kobayashi, and Turney into Kostoff to have created the claimed invention, It would have been obvious and desirable to have combined the punctuation and stop word removal technique of Kirsch into Kostoff so that the documents passes would have been more efficient, It would have been obvious and desirable to have combined the synonym word replacement of

Kobayashi into Kostoff so that the word counts could have been uniform across all of the documents, which would have yielded the most accurate clustering results, It would have been obvious and desirable to have combined the duplicate word removal of Turney into Kostoff so that the lists would have been uniform among all the documents in the cluster, This would have yielded the most accurate clustering results among the documents.

Regarding dependent claim 8, Kostoff teaches inputting one or more stop words, synonyms and a frequency level in col, 4 lines 39-49, col. 5 lines 59-64, and col. 6 lines 60-64.

Regarding dependent claim 9, Kostoff teaches adding words to a table in fig. 2, table 1, col. 4 lines 50-68, and col. 6 line 65- col. 7 line 11. Kostoff teaches determining the frequency of each word remaining n the table in fig. 2, table 1, col. 4 lines 50-68, and col. 6 line 65- col. 7 line 11. Kostoff teaches removing words below a frequency level from the table in col, 6 lines 2-64.

Kostoff does not teach removing punctuation and case from the documents, Kostoff does not teach removing stop words from the document, Kostoff does not teach replacing words in the documents with synonyms. Kostoff does not teach removing duplicate words from the documents, Kirsch teaches removing punctuation and case from the documents in col. 12 lines 5-7. Kirsch teaches removing stop words from the document in col. 12 lines 13-15. Kobayashi teaches replacing words in the documents with synonyms in fig. 3, 34-35, and col, 1 line 54 - col, 2 line 13. Turney teaches removing duplicate words from the documents in col. 5 lines 37-38.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to have combined Kirsch, Kobayashi, and Turney into

Kostoff to have created the claimed invention, It would have been obvious and desirable to have combined the punctuation and stop word removal technique of Kirsch into Kostoff so that the documents passes would have been more efficient, It would have been obvious and desirable to have combined the synonym word replacement of Kobayashi into Kostoff so that the word counts could have been uniform across all of the documents, which would have yielded the most accurate clustering results, It would have been obvious and desirable to have combined the duplicate word removal of Turney into Kostoff so that the lists would have been uniform among all the documents in the cluster, This would have yielded the most accurate clustering results among the documents,

Regarding dependent claim 10, Kostoff teaches inputting one or more stop words, synonyms and a frequency level in col. 4 lines 39-49, col, 5 lines 59-64, and col. 6 lines 60-64.

Regarding dependent claim 12, Kostoff teaches adding words to a dictionary table in fig. 2, table 1, col, 4 lines 50-68, and col, 6 line 65 - col. 7 line 11. Kostoff teaches determining the frequency of each word remaining the table in fig. 2, table 1, col, 4 lines 50, and col. 6 line 65 - col. 7 line 11. Kostoff teaches removing words below a frequency level from the dictionary table in col. 6 lines 2-64.

Kostoff does not teach removing punctuation and case from the documents, Kostoff does not teach removing stop words from the document, Kostoff does not teach replacing words in the documents with synonyms. Kostoff does not teach removing duplicate words from the documents, Kirsch teaches removing punctuation and case from the documents in col. 12 lines 5-7. Kirsch teaches removing stop words from the document in col, 12 lines 13-15. Kobayashi teaches replacing words in the documents with synonyms in fig. 3, 34-35, and col.

1 line 54 - col. 2 line 13. Turney teaches removing duplicate words from the documents in col, 5 lines 37-38.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to have combined Kirsch, Kobayashi, and Turney into Kostoff to have created the claimed invention. It would have been obvious and desirable to have combined the punctuation and stop word removal technique of Kirsch into Kostoff so that the documents passes would have been more efficient, It would have been obvious and desirable to have combined the synonym word replacement of Kobayashi into Kostoff so that the word counts could h ye been uniform across all of the documents, which would have yielded the most accurate clustering results. It would have been obvious and desirable to have combined the duplicate word removal of Turney into Kostoff so that the lists would have been uniform among all the documents in the cluster. This would have yielded the most accurate clustering results among the documents.

Regarding dependent claim 13, Kostoff teaches inputting one or more stop words, synonyms and a frequency level in col. 4 lines 39-49, col. 5 lines 59-64, and col, 6 lines 60-64.

Regarding dependent claim 14, Kostoff teaches adding words to a table in fig. 2, table 1, col. 4 lines 50-68, and col. 6 line 65- col, 7 line 11. Kostoff teaches determining the frequency of each word remaining n the table in fig. 2, table 1, col. 4 lines 50-68, and col. 6 line 65 - col. 7 line 11. Kostoff teaches removing words below a frequency level from the table in col, 6 lines 2-64.

Kostoff does not teach removing punctuation and case from the documents, Kostoff does not teach removing stop words from the document. Kostoff does not teach replacing words in the documents

with synonyms, Kostoff does not teach removing duplicate words from the documents, Kirsch teaches removing punctuation and case from the documents in col. 12 lines 5-7. Kirsch teaches removing stop words from the document in col, 12 lines 13-15. Kobayashi teaches replacing words in the documents with synonyms in fig. 3, 34-35, and col, 1 line 54 - col. 2 line 13. Turney teaches removing duplicate words from the documents in col, 5 lines 37-38.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to have combined Kirsch, Kobayashi, and Turney into Kostoff to have created the claimed invention. It would have been obvious and desirable to have combined the punctuation and stop word removal technique of Kirsch into Kostoff so that the documents passes would have been more efficient, It would have been obvious and desirable to have combined the synonym word replacement of Kobayashi into Kostoff so that the word counts could have been uniform across all of the documents, which would have yielded the most accurate clustering results. It would have been obvious and desirable to have combined the duplicate word removal
of Turney into Kostoff so that the lists would have been uniform among all the documents in the cluster. This would have yielded the most accurate clustering results among the documents.

Regarding dependent claim 15, Kostoff teaches inputting stop words in col. 4 lines 39-49, col. 5 lines 59-64, and col, 6 lines 60-64.

Regarding dependent claim 16, Kostoff teaches inputting synonyms in col. 4 lines 39-49, col. 5 lines 59-64, and col. 6 lines 60-64.

Regarding dependent claim 17, Kostoff teaches inputting a frequency level in col, 4 lines 39-49, col. 5 lines 59-64, and col. 6 lines 60-64.

2. Appellants' Position

a. Dependent Claims 2-5, 7-10, and 12-17

With respect to dependent claims 2-5, 7-10, and 12-17, the Office Action makes reference to the prior art Kirsch, Kobayashi, and Turney as teaching concepts such as removing punctuation, replacing words with synonyms, removing stop words, removing duplicates words, clustering, etc.

As discussed above, contrary to the highly manual process described in Kostoff, the claimed methodology defined by independent claims 1, 6, and 11 is fully automated (the only input required being the "maximum dictionary size", which can simply be equal to the available memory or manually preset by the user), while Kostoff requires the user to manually create the trivial phrase list (col. 4, lines 39-42). The efficiency gains of the automated inventive methodology when compared to the manual system described in Kostoff are substantial.

Further, the removal of trivial words is similar to the claimed removal of a manually created list of "stop" words (the, and, a, there, is, than) as defined by dependent claims 2-3, 7-8, and 12-13. The rules of claim differentiation and construction provide that each claim in a patent is presumptively different in scope. Therefore, the removal of trivial stop words in the dependent claims is different that the removal of words based on the maximum dictionary size in the independent claims. Here, the removal of a manually created list of trivial phrases ("to", "if", etc.) in Kostoff is equivalent to the claimed removal of a manually created list of stop words (the, and, a, there, is, than). Thus, the claimed method of limiting the dictionary according to a maximum size is a distinct feature from the removal of trivial or stop words and phrases. Therefore, it is Appellants' position that the discussion in Kostoff regarding the list of trivial words and phrases teaches no more that what is performed when the claimed invention removes stop words. There is nothing within Kostoff which would suggest that this removal of trivial or stop

words would lead one ordinarily skilled in the art to limit which words are to be included in the dictionary according to a "maximum dictionary size".

The creation of a manual list of trivial words ("to", "if", etc.) and its removal from the dictionary does not suggest the claimed automated methodology which simply and automatically limits the dictionary using a size limit. It is Appellants' position that the requirement that a manually created list be used to limit the dictionary size teaches away from the claimed automated methodology which does not require the user to specify any words, but instead merely eliminates the least frequent words from the dictionary. Further, the claimed invention may actually include all "trivial" words (if these stop words are not otherwise removed as provided in the dependent claims) as these words may be the most common. Again, the claimed invention removes the "most frequently occurring words in said documents as limited by said "maximum dictionary size"" and trivial or stop words may actually be the most common (if not removed).

Thus, dependent claims 2-5, 7-10, and 12-17 are similarly patentable, because of the additional features they define and because they depend from patentable independent claims. In view of the foregoing, the Board is respectfully requested to reconsider and withdraw this rejection.

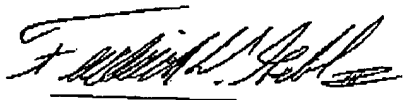
C. CONCLUSION

In view the forgoing, the Board is respectfully requested to reconsider and withdraw the rejections of claims 1-17.

Appeal Brief
10/320,318

Please charge any deficiencies and credit any overpayments to Attorney's Deposit Account Number 09-0441.

Respectfully submitted,



Frederick W. Gibb, III
Registration No. 37,629

Date:11/8/05
Gibb I.P. Law Firm, LLC
2568-A Riva Road, Suite 304
Annapolis, MD, 21401
301-261-8071
Customer No. 29154

VIII. CLAIMS APPENDIX

1. (Previously Presented) A method of automatically creating a dictionary for clustering text documents comprising:

inputting a maximum dictionary size;

determining a frequency of each word in each of said documents;

creating a dictionary of most frequently occurring words in said documents as limited by said maximum dictionary size, such that said dictionary contains less than all words in said documents;

after creating said dictionary, determining a frequency of phrases in each of said documents that contain only words in said dictionary;

adding most frequently occurring phrases to said dictionary; and

outputting said most frequently occurring words and said most frequently occurring phrases as said dictionary, wherein said dictionary size limits the number of words and phrases maintained in said dictionary.

2. (Previously Presented) The method in claim 1, wherein said determining a frequency of each word comprises:

removing punctuation and case from said documents;

removing stop words from said document;

replacing words in said documents with synonyms;

removing duplicate words from said documents;

adding remaining words to said dictionary as limited by said maximum dictionary size;

determining said frequency of each word remaining in said dictionary; and

removing words below a frequency level from said dictionary.

3. (Original) The method in claim 2, further comprising inputting one or more of said stop words, said synonyms, and said frequency level.

4. (Previously Presented) The method in claim 1, wherein said determining a frequency of phrases comprises:

removing punctuation and case from said documents;

removing stop words from said document;

replacing words in said documents with synonyms;

adding said phrases in each of said documents that contain only words in said dictionary to said dictionary;

determining said frequency of said phrases remaining in said dictionary; and

removing phrases below a frequency level from said dictionary.

5. (Original) The method in claim 4, further comprising inputting one or more of said stop words, said synonyms, and said frequency level.

6. (Previously Presented) A method of automatically creating a dictionary for clustering text documents comprising:

inputting a maximum dictionary size;

performing a first pass for each of said documents comprising:

determining a frequency of each word in each of said documents; and

creating a dictionary of most frequently occurring words in said documents as limited by said maximum dictionary size, such that said dictionary contains less than all words in said documents;

after performing said first pass, performing a second pass for each of said documents comprising:

determining a frequency of phrases in each of said documents that contain only words in said dictionary; and

adding most frequently occurring phrases to said dictionary; and

outputting said most frequently occurring words and said most frequently occurring phrases as said dictionary, wherein said dictionary size limits the number of words and phrases maintained in said dictionary.

7. (Previously Presented) The method in claim 6, wherein said determining a frequency of each word comprises:

removing punctuation and case from said documents;

removing stop words from said document;

replacing words in said documents with synonyms;

removing duplicate words from said documents;

adding remaining words to said dictionary as limited by said maximum dictionary size;

determining said frequency of each word remaining in said dictionary; and

removing words below a frequency level from said dictionary.

8. (Original) The method in claim 7, further comprising inputting one or more of said stop words, said synonyms, and said frequency level.

9. (Previously Presented) The method in claim 6, wherein said determining a frequency of phrases comprises:

removing punctuation and case from said documents;

removing stop words from said document;

replacing words in said documents with synonyms;

adding said phrases in each of said documents that contain only words in said dictionary to said dictionary;

determining said frequency of said phrases remaining in said dictionary; and

removing phrases below a frequency level from said dictionary.

10. (Original) The method in claim 9, further comprising inputting one or more of said stop words, said synonyms, and said frequency level.

11. (Previously Presented) A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform a method of automatically creating a dictionary for clustering text documents, said method comprising:

inputting a maximum dictionary size;

determining a frequency of each word in each of said documents;

creating a dictionary of most frequently occurring words in said documents as limited by said maximum dictionary size, such that said dictionary contains less than all words in said documents;

after creating said dictionary, determining a frequency of phrases in each of said documents that contain only words in said dictionary;

adding most frequently occurring phrases to said dictionary; and

outputting said most frequently occurring words and said most frequently occurring phrases as said dictionary, wherein said dictionary size limits the number of words and phrases maintained in said dictionary.

12. (Previously Presented) A program storage device as in claim 11, wherein said determining a frequency of each word comprises:

removing punctuation and case from said documents;

removing stop words from said document;

replacing words in said documents with synonyms;

removing duplicate words from said documents;

adding remaining words to said dictionary;

determining said frequency of each word remaining in said dictionary; and

removing words below a frequency level from said dictionary.

13. (Original) A program storage device as in claim 12, further comprising inputting one or more of said stop words, said synonyms, and said frequency level.

14. (Previously Presented) A program storage device as in claim 11, wherein said determining a frequency of phrases comprises:

removing punctuation and case from said documents;

removing stop words from said document;

replacing words in said documents with synonyms;

adding said phrases in each of said documents that contain only words in said dictionary to said dictionary;

determining said frequency of said phrases remaining in said dictionary; and

removing phrases below a frequency level from said dictionary.

15. (Original) A program storage device as in claim 14, further comprising inputting said stop words.

16. (Original) A program storage device as in claim 14, further comprising inputting said synonyms.

17. (Original) A program storage device as in claim 14, further comprising inputting said frequency level.

Appeal Brief
10/320,318

IX. EVIDENCE APPENDIX

There is no other evidence known to Appellants, Appellants' legal representative or Assignee which would directly affect or be directly affected by or have a bearing on the Board's decision in this appeal.

X. RELATED PROCEEDINGS APPENDIX

There is no other related proceeding known to Appellants, Appellants' legal representative or Assignee which would directly affect or be directly affected by or have a bearing on the Board's decision in this appeal.